

CLASSIFICAÇÃO DE IMAGENS ATRAVÉS DE CARACTERÍSTICAS DE COR E TÉCNICAS DE CLUSTERIZAÇÃO

Juliano Gomes Weber

Marcos Cordeiro d'Ornellas

julianors@gmail.com

marcosdornellas@gmail.com

Universidade Federal de Santa Maria (UFSM)

Centro de Tecnologia (CT) Predio 07 Sala 216 , 97105-900

Santa Maria, RS, Brasil

***Abstract.** This paper proposes a new method to extract features from images using a corner detection and clustering techniques. The goal is to store these features in a DBMS, and then use them to Content Based Image Retrieval. Besides encouraging results, the proposed method has a low computational complexity and with satisfactory results.*

1. Introdução

Os preços cada vez mais acessíveis de equipamentos que permitem a geração e armazenamento de imagens digitais, têm tornado as pesquisas em recuperação de imagens por conteúdo cada vez mais relevantes em nossos dias, devido à sistemas eficazes para gerenciar toda esta informação gerada. Aliado a esses fatores, o crescimento e popularização da *World Wide Web* torna a disponibilização de imagens digitais bastante abundante. Essas grandes coleções de imagens estão disponíveis nas mais diversas áreas do conhecimento, como: imagens médicas, presentes no trabalho de Antani et al. (2008) e Chang et al. (2008), sensoriamento remoto, conforme Tobin et al. (2006), Liu and He (2008) e Xie et al. (2008). A popularização do acesso a imagens digitais tem permitido que esta área de pesquisa, busque soluções específicas para a recuperação de imagens em setores da sociedade antes não atacados, como por exemplo organização de imagens digitais pessoais, o que pode ser observado em um recente trabalho proposto por Wen et al. (2008). No ano de 2006, Kompatsiaris (2006) produziu algumas estimativas que demonstraram alguns números impressionantes para a produção audiovisual nos anos vindouros. Em todo o mundo, 1-2 hexabytes (bilhões de gigabytes) de conteúdo eletrônico seriam produzidos e 80 bilhões de imagens digitais seriam produzidas. Além disso, mais de um bilhão de imagens relacionadas a transações comerciais estariam disponíveis e deveriam aumentar dez vezes nos dois anos seguintes.

1.1 Trabalhos relacionados

Os sistemas de recuperação de imagens por conteúdo primeiros datam da década de 80 Chang and Fu (1980), porém os mais importantes começam a surgir durante a década de 90. Dentre os sistemas comerciais de recuperação de imagens baseados em conteúdo mais conhecidos são o QBIC da IBM Flickner et al. (1995), que foi pioneiro na área e é um dos mais importantes até hoje e o Virage que foi desenvolvido para recuperação de vídeos e imagens. Após, outros sistemas foram desenvolvidos, como o Netra ying Ma and Manjunath (1999) e o Photobook Pentland et al. (1996), onde foram sendo agregados novos recursos que incluíram características de cor e textura para descrever seu conteúdo. O Blobworld Carson et al. (1999)

introduziu o uso de segmentação de imagens e o PicHunter implementou um mecanismo de realimentação de relevância (*relevance feedback*). O MARS Rui et al. (1997) também propôs uma solução de realimentação de relevância e tem destaque perante os demais sistemas pelo envolvimento de múltiplas pesquisas em sua implementação, que incluem visualização computacional, recuperação de informação e sistemas de gerenciamento de base de dados.

Atualmente encontram-se disponíveis via Web, sistemas de domínio público como o GIFT (*GNU Image-Finding Tool*) e várias versões de demonstrações disponíveis on-line, como o Viper e o Compass (*Computer Aided Search System*)

Uma descrição mais apurada desses e de outros sistemas CBIR (*Content Based Image Retrieval*) pode ser vista em Veltkamp and Tanase (2002). Tendências das pesquisas selecionadas a sistemas CBIR pode ser visto em Datta et al. (2008).

Recentemente, Venugopal and Sudhamani (2008) utilizou uma técnica de agrupamento de dados em conjunto com uma técnica de indexação R* Tree em um sistema CBIR. Devido a semelhança deste trabalho presente no estado da arte da literatura, com o método proposto na pesquisa aqui apresentada, os resultados de Venugopal and Sudhamani (2008) foram utilizados como parâmetro de comparação para atestar a eficiência do método desenvolvido.

2. Materiais e métodos

A base de testes utilizada foi a de denominação *groundtruth*, disponível em <http://www.cs.washington.edu/research/imagedatabase/groundtruth>. Esta base possui 1085 imagens pertencentes à 20 categorias variadas. Todas estas imagens possuem o mesmo tamanho, 756 x 504 pixels, foram pré-processadas e seus metadados armazenados em um banco de dados.

A proposta deste trabalho é criação um novo método que utiliza de maneira apropriada duas técnicas da área da Visão Computacional presentes no estado da arte da literatura, que originalmente possuem objetivos distintos, para que atuando em conjunto consigam gerar metadados robustos, que posteriormente são utilizados eficientemente em um Sistema de Recuperação de Imagens por Conteúdo - CBIR. A primeira técnica se refere à detecção de cantos proposto por Harris and Stephens (1988) e que está presente em recentes publicações á área da visão Computacional, como por exemplo, Gevrekci and Gunturk (2009), van de Weijer et al. (2005), e é apontada por Zou et al. (2008) como a mais robusta técnica disponível com este objetivo específico. A segunda é a técnica de agrupamento (*mean-shift*), uma técnica também presente no estado da arte na área de visão computacional que pode ser visto em Zhao et al. (2008), Kalpathy-Cramer and Hersh (2008), Shotton et al. (2008), Venugopal and Sudhamani (2008), Li et al. (2004), e que possui características primárias que atendem plenamente o problema a ser resolvido por este trabalho, que é a classificação de imagens coloridas.

O método (*mean-shift*), além de gerar agrupamentos com bastante qualidade, não exige um conhecimento prévio do número de agrupamentos a ser gerado, o que torna o método bastante viável para este trabalho, uma vez que o objetivo é classificar imagens coloridas heterogêneas, onde nunca se terá conhecimento prévio do "possível melhor" número de agrupamentos ser gerado. Com o objetivo de otimizar a ação do algoritmo Harris foi empregado uma técnica de baixa complexidade computacional para corrigir a iluminação das imagens durante a fase de pré-processamento. A correção da iluminação tem o objetivo de otimizar a ação do algoritmo Harris Harris and Stephens (1988), evitando que sejam detectados falsos positivos, como por exemplo cantos em sombras, bem como, realçar cantos que poderiam não ser detectados na imagem original.

2.1 Arquitetura

Na Figura 2 observa-se o que se pode chamar de duas grandes fases do método desenvolvido, onde há uma fase de pré-processamento, que ocorre de modo *off-line* e uma fase de indexação, classificação e recuperação das imagens.

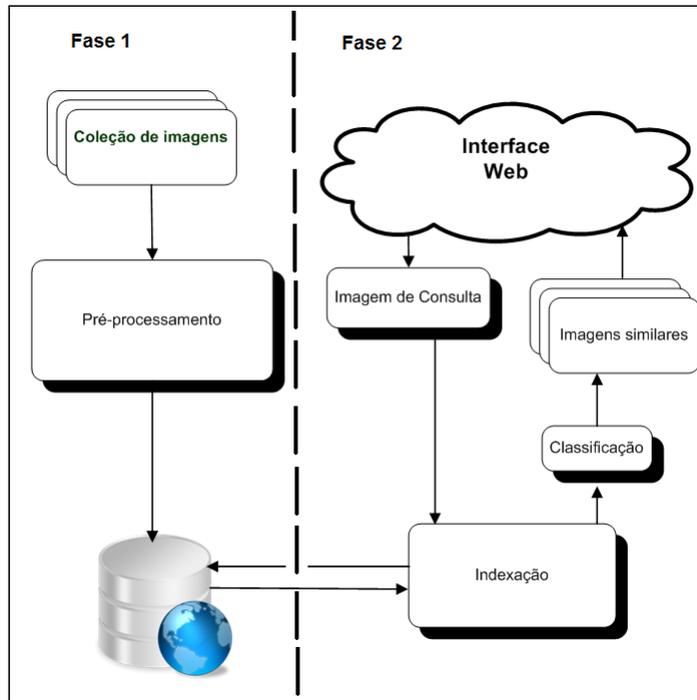


Figure 1: Arquitetura do método desenvolvido

2.2 Pré-processamento

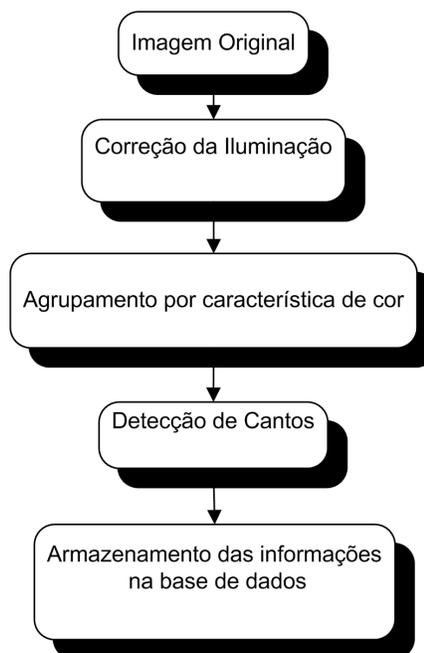


Figure 2: Fases do pré-processamento empregado pelo método desenvolvido

O pré-processamento desenvolvido neste trabalho objetiva eliminar a produção de agrupamentos fragmentados, ou seja, evitar a que o agrupamento da imagem seja efetuado várias vezes durante o processo de comparação das imagens, o que traria um acréscimo bastante elevado da complexidade computacional do método. Com o uso do pré-processamento, o método desenvolvido armazena uma vez só na base de dados todos os metadados necessários para a tarefa de indexação e comparação de imagens, baseadas em suas regiões e seus cantos detectados, ou seja, o método aplica o algoritmo de agrupamento e detecção de cantos somente uma vez durante todo o fluxo de execução do método. Todo o pré-processamento é efetuado off-line. Os metadados extraídos durante este processo são armazenados em um banco de dados relacional. A Figura 2 ilustra em linhas gerais os passos executados pelo método na fase de pré-processamento, enquanto a Figura 3 apresenta a GUI (*Graphical user interface*) desenvolvida para facilitar este procedimento.

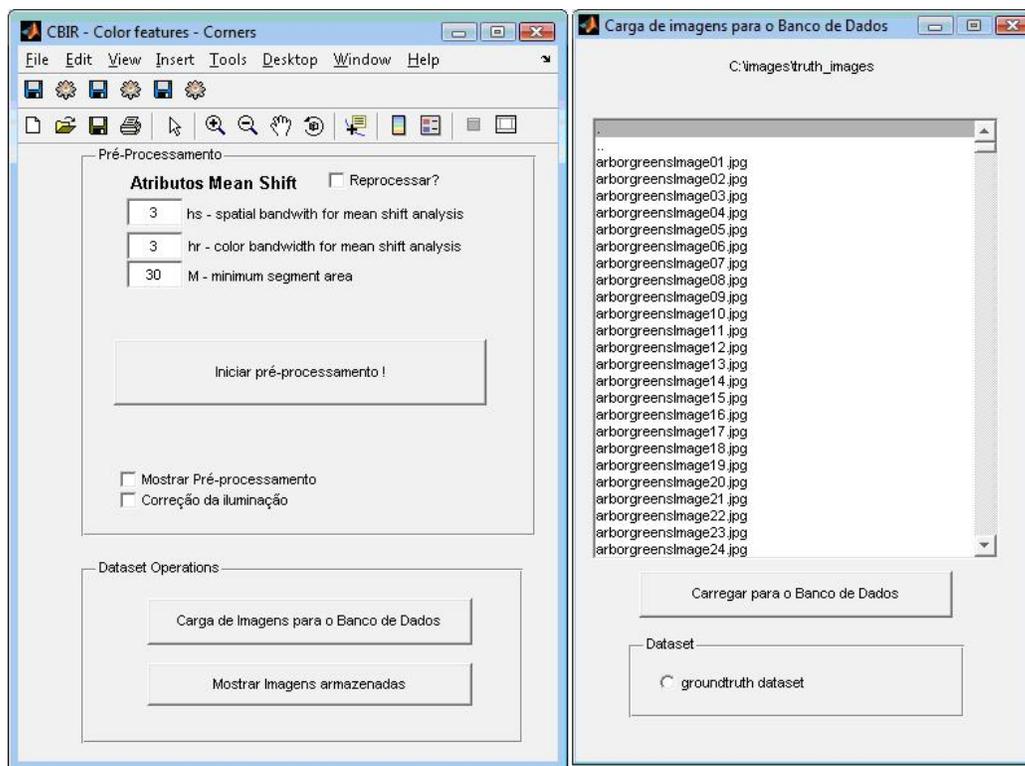


Figure 3: - Interface desenvolvida em MATLAB utilizada na fase de pré-processamento do método desenvolvido

A fase de agrupamento de dados é de fundamental importância na solução proposta por este trabalho. Desta forma, foram efetuados vários experimentos para o aperfeiçoamento da fase de segmentação, tentando chegar a um bom nível de segmentação, porém com um tempo de processamento baixo. Considerando que a segmentação ocupa em média, cerca de 60% do tempo total do processamento das imagens, esta análise se tornou importante, para que obtivéssemos bons resultados com um baixo tempo de processamento. O algoritmo *mean-shift* permite definir três parâmetros de entrada, que são:

Spatial bandwidth: é um valor que define o tamanho da janela de pesquisa na imagem, e por isso, recebe valores positivos inteiros.

Color bandwidth: Este parâmetro define o valor limite da distância entre duas cores, a partir do qual elas são tomadas como diferentes, ou pertencendo a conjuntos diferentes.

Minimum segment area: a área mínima das regiões obriga a que cada região tenha uma área

igual ou superior a esse valor e toma também números inteiros positivos.

Para verificar quais valores de parâmetros produziam melhores resultados, foram efetuados testes com 5 imagens de dimensão 756 x 504 pixels, variando os valores dos parâmetros. A primeira observação é que o parâmetro *color bandwidth* não afeta de maneira importante o tempo de processamento, sendo que a diferença no tempo de processamento alterando os valores deste parâmetro, fica na casa de centésimos de segundos. Por outro lado, o parâmetro *spatial bandwidth* afeta bastante o tempo de processamento. Considerando que este trabalho aplicaria o método de detecção de cantos na imagem agrupada por características de cor, durante os testes foi considerado o número de cantos detectados em função da variação dos parâmetros do método *mean-shift*. A conclusão foi que, na maioria dos casos, os melhores resultados foram obtidos fixando os parâmetros, durante a fase de pré-processamento com os seguintes valores: *Spatial bandwidth* = 3; *Color bandwidth* = 3; *Minimum segment area* = 30; A tabela 1 apresenta os testes considerando a variação do parâmetro *Spatial bandwidth* e seus respectivos tempos médios de segmentação por imagem.

Correção da Iluminação É conhecido que todo corretor de iluminação realça o que muitas vezes o que antes não aparecia na imagem, bem como elimina sombras indesejadas da mesma. Com o objetivo de aperfeiçoar a ação do detector de cantos empregado neste trabalho, no sentido de não considerar sombras da imagem como cantos e também detectar cantos que na imagem original não seriam detectados, foi empregada uma fase de correção de iluminação durante o pré-processamento das imagens. Esta técnica foi proposta por Jahne (1999), baseada em operações básicas da morfologia matemática e possui baixa complexidade computacional e que atendeu perfeitamente as necessidades deste trabalho.



(a) Imagem original e agrupada com os cantos detectados



(b) Imagem com iluminação otimizada, agrupada com os cantos detectados

Figure 4: Correção da iluminação

Na Figura 4(a) está um exemplo da aplicação do algoritmo de detecção de cantos na imagem agrupada original, ou seja, sem a correção de iluminação aplicada, já Figura 4(b) foi aplicada a detecção de cantos na imagem agrupada com a correção da iluminação. É perceptível que os cantos detectados na imagem representada pela Figura 4(b) possuem cantos detectados com maior precisão, por exemplo, basta observar as "sombras" que na Figura 4(a) foram consideradas como canto, já não imagem representada pela Figura 4(b), não. Outro aspecto importante é a detecção de cantos reais, o que pode ser observada nos pilares dos prédios, que de fato são cantos da imagem, e não haviam sido detectados na Figura 4(b).

Agrupamento O método de agrupamento utilizado por este trabalho trata-se de um método iterativo que estima a moda (regiões mais densas) de distribuições multivariadas obtidas no espaço de atributos. A quantidade de agrupamentos é obtida automaticamente pela descoberta dos centros destas modas. Uma vez que o objetivo deste trabalho é a classificação de imagens coloridas, o método *mean-shift* se faz ideal, uma vez que nunca será conhecido previamente o número de agrupamentos gerados em uma imagem colorida, quando se faz agrupamento por característica de cor em imagens heterogêneas. Após a aplicação deste método, apenas algumas cores se mantêm na imagem resultado, e é possível extrair informações como número de regiões, cantos destas regiões, tamanho e circunferências destas regiões. Este trabalho priorizou a extração dos cantos das regiões como principal metadado a ser utilizado na comparação entre as imagens. A

Deteção de Cantos Este trabalho aplica o algoritmo de deteção de cantos desenvolvido por Harris and Stephens (1988) na imagem agrupada por característica de cor e com iluminação corrigida. A escolha desta técnica se deve ao fato deste método estar presente no estado da arte da literatura, com superioridade comprovada, que pode ser visto em recente publicação de Zou et al. (2008). A Figura 5 demonstra o resultado da aplicação do método de deteção de cantos na imagem agrupada e com sua intensidade de iluminação corrigida, porém apresentados na imagem original, o que atesta a robustez da deteção de cantos, onde mesmo aplicando-o em uma imagem com menor complexidade computacional, os cantos detectados são compatíveis. O foco deste trabalho é tornar os cantos das regiões geradas pelo método *mean-shift*, o principal metadado da imagem.



Figure 5: Cantos detectados na imagem agrupada e apresentados na imagem original

Indexação de Imagens Os passos efetuados durante o processo de indexação dos metadados extraídos, visando sua utilização posterior na recuperação das imagens por semelhança estão dispostos na Figura 6.

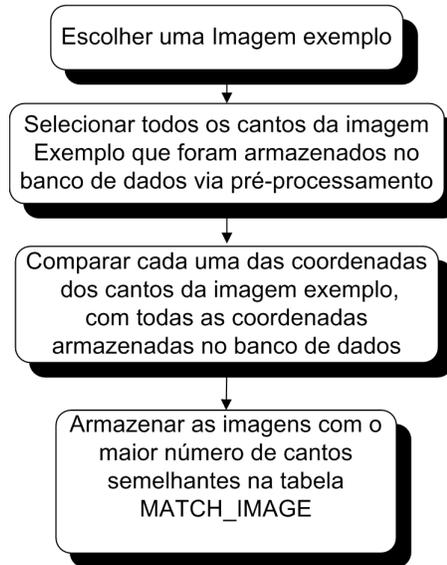


Figure 6: Procedimento de indexação das imagens

A tabela 2 demonstra a instrução SQL responsável pela indexação dos metadados extraídos pelo método desenvolvido. O objetivo é armazenar os quantificadores necessários na tabela MATCH_IMAGE, que pode ser visualizado na Figura 7.

Esta operação é efetuada via uma instrução SQL. A formulação matemática para o critério de indexação está definida pela equação (1).

$$I(x_i, y_i) \begin{cases} 1, & \text{if } I(x_i, y_i) \leq K(x_i - \sigma, y_i - \sigma) \leq K(x_i + \sigma, y_i + \sigma) \\ 0, & \text{if } I(x_i, y_i) > K(x_i - \sigma, y_i - \sigma) > K(x_i + \sigma, y_i + \sigma) \end{cases} \quad (1)$$

Para $i = 1$ até n , onde n é o número de corners armazenados via pré-processamento, e I armazena os corners de todas as imagens armazenadas na base de dados e K são os corners da imagem exemplo os quais estão sendo comparados para verificar sua similaridade. Assim sendo, a imagem mais semelhante será a que retornar 1 o maior número de vezes, durante a execução da indexação. Objetivando um maior dinamismo na comparação entre as imagens, foi, adicionalmente, criado um fator de variação σ que permite que cantos detectados em áreas próximas (pixels vizinhos) aos cantos detectados originalmente na fase de pré-processamento, sejam também considerados como similares aos cantos da imagem exemplo. Foram efetuados testes variando σ de 1 a 30, e os resultados de eficiência armazenados. Na média, melhores resultados apresentados foram com σ com valor 20.

Table 1: Instrução SQL responsável pela classificação das imagens por ordem de semelhança

```

SELECT a.id_image, count(*), a.coord_x,a.coord_y,c.coord_x,c.coord_y
FROM corner a, image b , corner_origem c
WHERE a.id_image = b.id_image and
(a.COORD_X >= c.coord_x -  $\sigma$  AND a.COORD_X <= c.coord_x +  $\sigma$  ) AND
(a.COORD_y >= c.coord_y -  $\sigma$  AND a.COORD_y <= c.coord_y +  $\sigma$  )
GROUP BY a.id_image order by 2 desc
  
```

O conjunto de dados retornados pelo processamento da consulta SQL demonstrada na tabela 2, é armazenado em uma tabela denominada MATCH_IMAGE, que será utilizada na fase de recuperação das imagens.

MATCH_IMAGE
ID_MATCH_IMAGE: Int;
ID_IMAGE_IN: Int;
ID_SIMILAR_IMAGE: Int;
NUMBER_CORNER_POINTS: Int;

Figure 7: Projeto físico da tabela MATCH_IMAGE

3. Resultados experimentais

protótipo computacional desenvolvido neste trabalho permite que após o pré-processamento do conjunto de imagens, os seus metadados sejam disponibilizados para qualquer tipo de aplicação que tenha possibilidade de acessar os dados em um banco de dados relacional. Neste contexto, os aplicativos Web, por sua natureza Cliente-Servidor conseguem utilizar-se da técnica desenvolvida neste trabalho com grande facilidade, tornando o protótipo computacional aqui desenvolvido, bastante versátil enquanto disponibilização da ferramenta para os mais diversos conjuntos de imagens possíveis.

Além da GUI (*Graphic User Interface*) desenvolvida para a fase de pré-processamento, foi desenvolvida uma interface Web, que permite a utilização do método implementado neste trabalho. A Figura 8 demonstra a interface Web desenvolvida.



Figure 8: Tela principal da Interface Web Desenvolvida

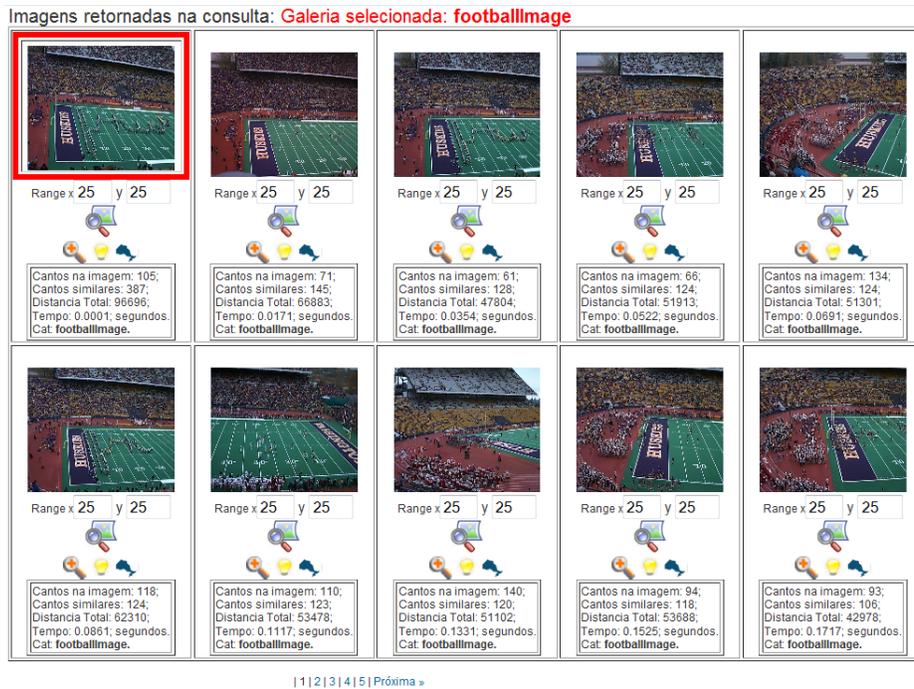


Figure 9: Exemplo de retorno de consulta por semelhança

4. Análise dos Resultados

Este trabalho comparou seu desempenho com os resultados apresentados pela recente publicação de Venugopal and Sudhamani (2008), que também utiliza o método de agrupamento imagens *mean-shift* como base no processo de extração de metadados das imagens, e por isto foi escolhido como ponto de comparação com os resultados deste trabalho.

4.1 Resultados sob o ponto de vista da complexidade computacional

Para verificar a complexidade computacional do método de indexação desenvolvido e também validar a eficiência dos resultados, o procedimento de testes ocorreu da seguinte forma: Foram eleitas aleatoriamente 70 imagens (não repetidas) como imagens exemplo, e a média dos resultados obtidos estão apresentadas na Figura 10.

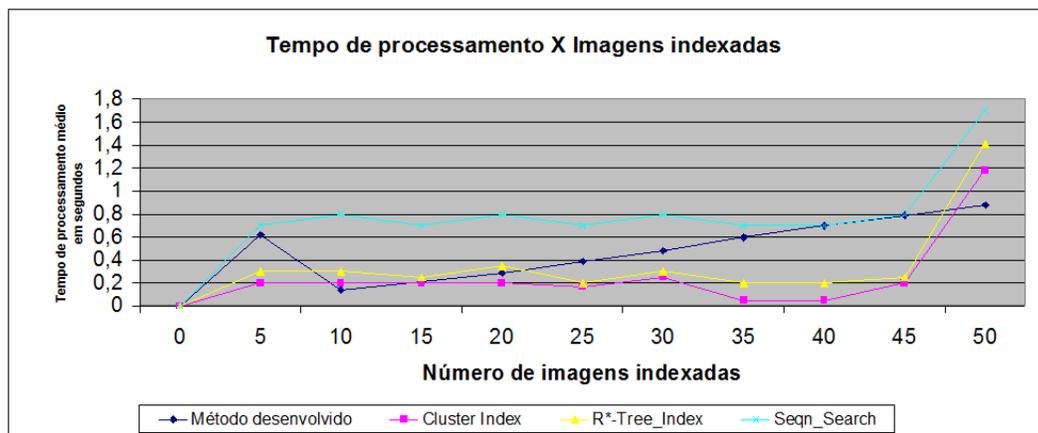


Figure 10: Resultados referentes à complexidade computacional do método desenvolvido

4.2 Resultados sob o ponto de vista da precisão dos resultados

O critério de avaliação para a precisão dos retornos obtidos após a execução do método está disposto a seguir:

$\text{precisao}(k) = c_k = k$; $\text{recall}(k) = c_k = M$; onde:

k = número de imagens retornadas;

c_k = número de retornos relevantes entre os k retornos;

M = Total de imagens relevantes encontradas no banco de dados.

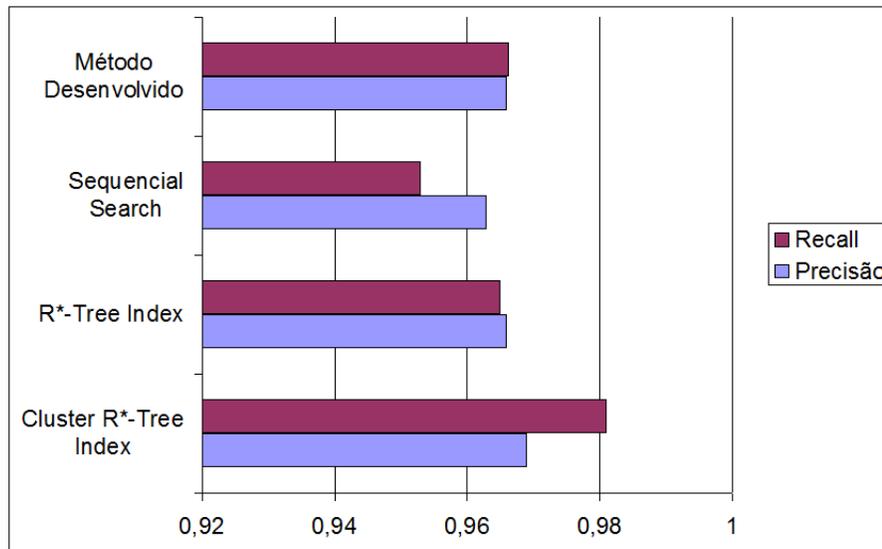


Figure 11: Representação gráfica dos resultados obtidos nos testes computacionais

5. Comentários Finais

5.1 Conclusões

Observa-se através dos experimentos efetuados e a comparação com métodos atuais de Venugopal and Sudhamani (2008), que os resultados são semelhantes no ponto de vista da precisão dos resultados, e superiores em complexidade computacional quando o número de imagens ultrapassa o número de 45 imagens indexadas.

Quanto à eficácia do método no que se refere ao número e qualidade de imagens retornadas por semelhança, também se percebe que os resultados são compatíveis com os resultados presentes nas publicações com mesmo fim, e com uso da mesma técnica de agrupamento de dados. Outro fator importante e que pode ser observado, é que em todos os testes efetuados, a imagem com maior *rank* retornada é a própria imagem de entrada (que está sempre incluída no conjunto de imagens pré-processadas), o que assegura a robustez do método desenvolvido.

REFERENCES

- Antani, S., Long, L. R., & Thoma, G. R., 2008. Bridging the gap: Enabling cbir in medical applications. *Computer-Based Medical Systems, IEEE Symposium on*, vol. 0, pp. 4–6.
- Carson, C., Belongie, S., Greenspan, H., & Malik, J., 1999. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 1026–1038.

- Chang, N.-S. & Fu, K.-S., 1980. Query-by-pictorial-example. *IEEE Trans. Softw. Eng.*, vol. 6, n. 6, pp. 519–524.
- Chang, Y., Antani, S., Lee, D. J., Gledhill, K., Long, L. R., & Christensen, P., 2008. Cbir of spine x-ray images on inter-vertebral disc space and shape profiles. In *CBMS '08: Proceedings of the 2008 21st IEEE International Symposium on Computer-Based Medical Systems*, pp. 224–229, Washington, DC, USA. IEEE Computer Society.
- Datta, R., Joshi, D., Li, J., & Wang, J. Z., 2008. Image retrieval: Ideas, influences, and trends of the new age. *ACM Comput. Surv.*, vol. 40, n. 2, pp. 1–60.
- Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., & Yanker, P., 1995. Query by image and video content: The qbic system. *Computer*, vol. 28, n. 9, pp. 23–32.
- Gevrekci, M. & Gunturk, B., 2009. Illumination robust interest point detection. vol. 113, n. 4, pp. 565–571.
- Harris, C. & Stephens, M., 1988. A combined corner and edge detector. In *The Fourth Alvey Vision Conference*, pp. 147–151.
- Jahne, B., 1999. *Handbook of Computer Vision and Applications: Volume 2: From Images to Features*. Academic Press, Inc., Orlando, FL, USA.
- Kalpathy-Cramer, J. & Hersh, W., 2008. Effectiveness of global features for automatic medical image classification and retrieval - the experiences of ohsu at imageclefmed. *Pattern Recogn. Lett.*, vol. 29, n. 15, pp. 2032–2038.
- Kompatsiaris, Y., 2006. Multimedia semantic analysis technologies and their potential uses. 2006. Disponível em http://www.samt2006.org/presentationsITI_MMAnalysis.pdf Acesso em 2009.
- Li, C.-Y., Shih, M.-C., & Hsu, C.-T., 2004. Image retrieval with relevance feedback based on graph-theoretic region correspondence estimation. In *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 3*, pp. 842–845, Washington, DC, USA. IEEE Computer Society.
- Liu, H. & He, G., 2008. Shape feature extraction of high resolution remote sensing image based on susan and moment invariant. In *CISP '08: Proceedings of the 2008 Congress on Image and Signal Processing, Vol. 2*, pp. 801–807, Washington, DC, USA. IEEE Computer Society.
- Pentland, A., Picard, R., & Sclaroff, S., 1996. Photobook: Content-based manipulation of image databases.
- Rui, Y., Huang, T. S., & Mehrotra, S., 1997. Content-based image retrieval with relevance feedback in mars. In *In Proc. IEEE Int. Conf. on Image Proc.*, pp. 815–818.
- Shotton, J., Blake, A., & Cipolla, R., 2008. Multiscale categorical object recognition using contour fragments. vol. 30, n. 7, pp. 1270–1281.
- Tobin, K., Bhaduri, B., Bright, E., Cheriyyadat, A., Karnowski, T., Palathingal, P., Potok, T., & Price, J., 2006. Automated feature generation in large-scale geospatial libraries for content-based indexing. vol. 72, n. 5, pp. 531–540.

- van de Weijer, J., Gevers, T., & Geusebroek, J.-M., 2005. Edge and corner detection by photometric quasi-invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, n. 4, pp. 625–630.
- Veltkamp, R. C. & Tanase, M., 2002. Content-based image retrieval systems: A survey. Technical Report TR UU-CS-2000-34 (revised version), Department of Computing Science, Utrecht University.
- Venugopal, C. R. & Sudhamani, M. V., 2008. Image retrieval from databases: an approach using region color and indexing technique. In *HPCNCS*, pp. 82–89.
- Wen, F., Chen, S., & Tang, X., 2008. Easytoon: cartoon personalization using face photos. In *MM '08: Proceeding of the 16th ACM international conference on Multimedia*, pp. 1021–1022, New York, NY, USA. ACM.
- Xie, Z., Roberts, C., & Johnson, B., 2008. Object-based target search using remotely sensed data: A case study in detecting invasive exotic australian pine in south florida. *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 63, n. 6, pp. 647 – 660.
- ying Ma, W. & Manjunath, B., 1999. Netra: A toolbox for navigating large image databases. In *Multimedia Systems*, pp. 568–571.
- Zhao, C., Knight, A., & Reid, I., 2008. Target tracking using mean-shift and affine structure. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–5.
- Zou, L.-h., Chen, J., Zhang, J., & Dou, L.-h., 2008. The comparison of two typical corner detection algorithms. In *IITA '08: Proceedings of the 2008 Second International Symposium on Intelligent Information Technology Application*, pp. 211–215, Washington, DC, USA. IEEE Computer Society.